Computation and Intentional Psychology

Michael Rescorla

Abstract: The *formal conception of psychological processes* (FCP) holds that psychological processes are not counterfactually sensitive to semantic properties of mental representations. FCP's opponents commonly object that it is incompatible with robust intentional psychological laws. Some of FCP's proponents respond that we do not need such laws. Others respond that the apparent incompatibility is illusory. I rebut the first response by articulating a rationale for scientific psychology to isolate intentional laws. I then argue that any proposed reconciliation of FCP and intentional explanation faces tenacious difficulties. My conclusion: FCP precludes a scientific psychology that achieves our explanatory ends.

§1. Are psychological processes formal?

The computational theory of mind holds that psychological processes are computational. *The formal conception of computation* holds that computational processes are formal manipulations of symbols, where "formal" means that the processes are not counterfactually sensitive to the meanings, contents, or semantic properties of symbols they manipulate. Block elucidates this counterfactual insensitivity, without endorsing it, as follows: "a given syntactic object would have caused the same output even if it had meant something quite different from what it actually means or even if it had meant nothing at all" (1990, p. 41). The computational theory of mind and the formal conception of computation jointly entail

> **The formal conception of psychological processes (FCP)**: Psychological processes are formal manipulations of symbols.

FCP is quite popular. Advocates include Egan (1992), Fodor (1994), Newell and Simon (1976), Pylyshyn (1984), Stich (1983), and many others. However, as critics often note, FCP seemingly conflicts with

> **The intentional laws thesis (ILT):** Scientific psychology should subsume mental states under laws that advert to their intentional contents.

If psychological processes are not sensitive to content, how can content play a non-trivial role in psychological laws?

Fodor has labored to answer this question for several decades. He advocates a theory of cognition based upon

> **The representational theory of mind (RTM):** Propositional attitudes are relations to mental representations, which have content. Attitudes inherit their contents from associated mental representations.

Fodor describes mental representations as comprising *the language of thought*, or *Mentalese*. He holds that psychological processes are sensitive to syntactic but not semantic properties of Mentalese: "[i]f two mental representations are identical in respect of… [their intrinsic, formal, nonrelational, nonsemantic] properties, then they play the same role in mental processes, even if their semantical properties (their truth-conditions, for example) are different" (Fodor, 1991a, p. 298). Thus, while Mentalese symbols have content, their content does not inform how the cognitive system manipulates them. Nevertheless, Fodor insists that psychology should isolate intentional laws. How, as Stich (1983, p. 96) inquires, can Fodor have it both ways?[1]

I will critique Fodor's (1994) most sustained attempt at reconciling FCP and ILT, an attempt subsequently elaborated by Rupert (2008) and Schneider (2005). I will argue that the difficulties facing Fodor are general and tenacious. My conclusion: FCP precludes an intentional

psychology that achieves our explanatory ends. Although this conclusion is familiar, many of my arguments differ from those commonly found in the literature.[2]

Since ILT mentions "laws," I must briefly elucidate my use of this phrase. How to distinguish between "lawlike" and "accidental" generalizations is a vexed question (Woodward, 2003, pp. 239-314). I assume that laws are counterfactual-supporting generalizations. I furthermore assume, following Fodor (1991b), that laws of the special sciences typically include either explicit or implicit *ceteris paribus* clauses. Some philosophers maintain that these clauses render the putative laws vacuous or untestable (Earman, Roberts, Smith, 2002). I join Fodor (1991b), Lange (2002), and Smith (2006) in holding that such worries can be answered satisfactorily. Other philosophers concede that *ceteris paribus* generalizations are non-vacuous but deny that they deserve to be called "laws," on the grounds that genuine laws are exceptionless. Those who agree can translate my talk about "laws" into talk about "counterfactual-supporting generalizations."

§2. Two important consequences of FCP

Throughout my discussion, I will rely heavily upon two consequences of FCP. The core idea underlying both doctrines is that syntax does not determine semantics, meaning, or content. A weak statement runs as follows:

SEMANTIC BARRENNESS: Content does not supervene upon syntactic type. For any Mentalese syntactic type *t*, there are possible tokens of *t* that have different contents.

For instance, a string of *n* strokes inscribed on a Turing machine tape is an inherently meaningless syntactic item, subject to any interpretation we choose. Virtually all commentators accept SEMANTIC BARRENNESS, or something closely resembling it (Haugeland, 1985, p.

91), (Newell and Simon, 1976, p. 116), Piccinini (2008, p. 216), (Putnam, 1988, p. 21), (Pylyshyn, 1984, p. 44), (Stich, 1983, p. 207). Indeed, most commentators accept

> **SUPERVENIENCE FAILURE:** A Mentalese symbol's content does not supervene on syntactic properties of mental computation. We do not fix the symbol's content even when we offer a complete syntactic description of the cognitive system's states and the mechanical rules governing transitions between states.

SUPERVENIENCE FAILURE is stronger than SEMANTIC BARRENNESS, because a complete syntactic description of some computational system will cite many syntactic properties beyond syntactic type, such as how one syntactic type characteristically interacts with others. Explicit advocates of SUPERVENIENCE FAILURE include Dennett (1987, p. 61), Egan (1999, p. 181), Field (2001, p. 58, p. 76), and many others.

FCP entails SUPERVENIENCE FAILURE. To claim that syntax determines semantics is to claim that a change in semantics entails a change in syntactic properties, that is, a change in properties to which computation is sensitive. Thus, to deny SUPERVENIENCE FAILURE is to regard mental computation as counterfactually sensitive to meaning or content. Since SUPERVENIENCE FAILURE entails SEMANTIC BARRENNESS, it follows that FCP entails SEMANTIC BARRENNESS.

SUPERVENIENCE FAILURE is particularly compelling if we endorse *content externalism*: the contents of mental states do not supervene upon internal neurophysiology. Most commentators assume *internalism about mental syntax*: syntactic properties of mental representations supervene upon internal neurophysiology. Content externalism and syntactic internalism jointly entail SUPERVENIENCE FAILURE. However, SUPERVENIENCE FAILURE neither entails nor presupposes these two doctrines. It readily follows from many

theories that recognize only *narrow content*, i.e. content that supervenes upon internal neurophysiology. As Peacocke (1994, p. 306) observes, SUPERVENIENCE FAILURE even applies to *and*-gates, since we can offer a complete syntactic characterization of some *and*-gate without specifying whether the "0"s and "1"s denote numbers, truth-values, something else, or nothing at all. Peacocke concludes: "the principle that syntax cannot determine semantics holds for anything recognizable as content" (1994, p. 307).

Block rejects SUPERVENIENCE FAILURE. He argues that *and*-gates are unrepresentative of those computations underlying cognitive activity (1990, pp. 32-36, pp. 41-42). According to Block, we fix the meanings of mental states by providing a complete syntactic description of the cognitive system, construed broadly to include functional relations born by internal states to perceptual inputs, behavioral outputs, and one another. Thus, meaning supervenes upon syntactic description, broadly construed. From his denial of SUPERVENIENCE FAILURE, Block concludes that FCP is false.

The early paper (Fodor, 1981) espouses a more extreme position, centered on a denial of SEMANTIC BARRENNESS. Fodor (1981, pp. 226-227) defends what he calls *the formality condition*: psychological processes "apply to representations in virtue of (roughly) the syntax of the representations," so they "are specified without reference to such semantic properties of representations as, for example, truth, reference, and meaning." The formality condition and SUPERVENIENCE FAILURE jointly entail FCP: if we can change a symbol's content without changing any syntactic properties, and if mental processes apply in virtue of syntax, then we can change a symbol's content without altering its role in mental processes. The formality condition, taken on its own, does not entail FCP. In fact, Fodor (1981) denies FCP. He argues that "mental representations affect behavior in virtue of their content," so that "beliefs of different content *can*

have different behavioral effects" (1981, p. 240). To reconcile the formality condition with his denial of FCP, Fodor argues that "two thoughts can be distinct in content only if they can be identified with relations to formally distinct representations" (1981, p. 227). In other words, he denies SEMANTIC BARRENNESS. On this view, any change in content necessarily entails a change in syntactic type. FCP is false, because a change in content entails a change in properties ---- syntactic properties --- to which computation is sensitive.

With the exception of Aydede (2005, p. 200), Fodor's early opposition to SEMANTIC BARRENNESS has found few subsequent advocates, and he himself has abandoned it. In his later writings, Fodor explicitly endorses SUPERVENIENCE FAILURE (1991a, pp. 281-2, pp. 297-8), SEMANTIC BARRENNESS (1994, pp. 22-24), and FCP (1991a, p. 298). This paper addresses Fodor's mature (1994) view, not his earlier (1981) view nor the views of Aydede (2005) and Block (1990). The latter three views are not instances of FCP.

Once we accept FCP, the obstacles to satisfactory intentional explanation become quite daunting. My goal in this paper is to highlight some particularly serious difficulties. I will first argue that scientific psychology should assign intentional laws a central explanatory role. I will then argue that FCP bars intentional laws from occupying that role.

§3. A rationale for intentional explanation

Stich (1983) accepts FCP and rejects ILT. He maintains that psychology should jettison content, confining itself to theories that cite syntactic, non-semantic properties of mental states. On the resulting *syntactic theory of mind*, cognitive science eschews intentional explanation, operating at a purely syntactic level. Why not settle for the syntactic theory of mind? What benefits accrue from psychological laws that cite intentional content? Although the literature

suggests various answers to this question, I will focus on a single fundamental consideration: we want to explain bodily movements under environment-involving descriptions.

A description is *environment-independent* if it applies to a neurophysiological duplicate of any creature to which it applies. Otherwise, the description is *environment-involving*. We frequently describe bodily movements in environment-involving terms. For instance, we might describe a certain bodily movement as: grasping a peanut; grasping a brown object in front of me; removing the only object from a table; and so on. Depending on the details of the case, we wish to explain some but not all of these environment-involving properties. In contrast, Stich claims that environment-involving descriptions are irrelevant to psychological explanation. He endorses *the autonomy principle*: "the states and processes that ought to be of concern to the psychologist are those that supervene on the current, internal, physical state of the organism" (1983, p. 164).[3] Following Burge (2007, pp. 226-228, pp. 335-336), Harman (1999, p. 240), Hornsby (1986), Peacocke (1993), and Pylyshyn (1984, pp. 23-38), I will assume that the autonomy principle is false. Specifically, I will assume that scientific psychology seeks to explain bodily movements under appropriate environment-involving descriptions.[4] Readers who disagree can regard my discussion as exploring this assumption's consequences.

How should we explain bodily motions under appropriate environment-involving descriptions? A natural strategy is to isolate *ceteris paribus* laws describing how mental states with certain contents induce a bodily motion with certain environment-involving properties. I will now argue that this is the *only* promising strategy. While many would agree with my assessment, I think that the reasoning behind it remains underappreciated.

Let us first ask whether the desired explanations ensue from a syntactic theory like that favored by Stich. Psychological laws fall into three categories, depending on whether they

describe relations between thoughts, relations between perceptual inputs and thoughts, or relations between thoughts and behavioral outputs. Following Devitt (1991), label these three categories T-T, I-T, and T-O. By assumption, T-T laws depict relations between mental states under syntactic descriptions. What about I-T and T-O laws? There are two promising options:

(1) I-T and T-O laws describe how mental states under syntactic descriptions relate to perceptual inputs and behavioral outputs under *environment-independent* physical/neurophysiological descriptions.

(2) I-T and T-O laws describe how mental states under syntactic descriptions relate to perceptual inputs and behavioral outputs under *environment-involving* physical/neurophysiological descriptions.

Fodor (1991a, pp. 281-282) endorses (1) as a template for computational psychology. Field (2001, pp. 73-74) endorses (2). He furthermore argues that a computational theory conforming to (2) can forego intentional explanation. I will now argue that neither (1) nor (2) promotes satisfactory explanations of bodily movements under environment-involving descriptions.

§3.1 Environment-independent T-O laws

Say we want to explain why John performs bodily motion *M*. In accord with (1), we specify relevant antecedent states, including John's propositional attitudes (described syntactically) and perceptual inputs (described environment-independently). Following Peacocke (1994), call these *the explaining states*. We also delineate I-T, T-T, and T-O laws dictating that, *ceteris paribus*, the explaining states induce *M* under an environment-independent description.

We do not thereby explain *M* under an environment-involving description. We do not even mention *M*'s environment-involving properties. Of course, *M* has such properties. For

instance, *M* might be a grasping of a peanut. What if we amend our theory by observing that *M* satisfies some environment-involving description? The explanatory gap persists. Explaining an event under description *s* and noting that the event satisfies description *t* is not the same as explaining the event under description *t* (Davidson, 1980, p. 154), (Peacocke, 1993), (Ruben, 2003, p. 199). If we explain why John performs a motor gesture and note that the gesture is a grasping of a peanut, we do not thereby explain why John grasps a peanut.[5]

What if we further supplement our theory with a semantics for Mentalese? Egan (1995, 1999) advocates this supplemented approach. On her conception, a good theory includes three elements: a syntactic theory of mental processes, including I-T and T-O laws conforming to (1); clauses stipulating that bodily motion *M* has some environment-involving property; and clauses stipulating that syntactic entities manipulated during computation have certain semantic properties. For instance, we might supplement a syntactic description of John's mental activity by observing that a Mentalese sentence in his "belief box" means *There is a peanut before me* while a sentence in his "desire box" means *I eat a peanut*.

It is difficult to see how Egan's supplementations advance our explanatory aims. Our theory treats semantic properties as irrelevant to how mental computation proceeds, because it declines to mention semantics within its laws. Semantic properties of explaining states do not bear upon which subsequent states ensue. If we begin with a theory that fails to explain bodily movements under environment-involving descriptions, how do we improve matters by stipulating irrelevant features of explaining states? Semantic properties contribute no added explanatory force.

§3.2 Environment-involving T-O laws

Suppose that Mentalese sentence *S* appears in John's desire box. By SEMANTIC BARRENNESS, syntactic type leaves meaning undetermined. Perhaps some tokens of *S* mean *I eat a peanut* while others mean *I drink a martini*. Depending on what *S* means, its appearance in the desire box tends to induce different environment-involving consequences. One set of typical consequences will ensue if *S* means *I eat a peanut*, another if *S* means *I drink a martini*, and so on. Uninterpreted syntactic description provides little guidance regarding a mental state's significance for environment-involving action. As Devitt puts it, "[a] thought has a distinctive role in producing behavior…. Syntax alone cannot explain that distinctive role" (1989, p. 382).

I do not claim that there are *no* T-O laws relating mental states under syntactic descriptions and bodily movements under environment-involving descriptions. I claim only that there are not *enough* T-O laws to serve our explanatory aims. For instance, it is plausible that we could build a robot instantiating a T-O law of this form: if the robot stands in certain relations to certain symbols, then *ceteris paribus* the robot moves five feet forward. The description "moves five feet forward" is environment-involving. Thus, (2) may help us explain bodily movements under certain environment-involving descriptions. But (2) seems unlikely to accommodate the full range of desired descriptions, which includes locutions such as "grasping a peanut."

We might modify (2) by allowing T-O laws that describe mental states under environment-involving, non-semantic descriptions. Our supplementary descriptions might mention John's causal-covariational relations to the external environment, his evolutionary history, features of his normal environment, and so on. (2) already taxonomizes perceptual inputs in this way. For instance, (2) allows us to describe a retinal stimulation as caused by a peanut. Suppose we extend (2) by allowing T-O laws that describe mental states under supplemented

descriptions like "bearing certain computational relations to a retinal stimulation caused by a peanut." Might such a law help us explain why John reaches for a peanut?

This suggestion faces a dilemma. (Cf. Peacocke, 1993, pp. 208-209.) Either our supplemented description of mental states does not determine their contents, or it does. If it does not, then the problem facing (2) persists. Depending on what *S* means, John's bearing some relation to *S* tends to yield quite different environment-involving consequences. So our theory is likely to leave unexplained certain relevant features of John's behavior. On the other hand, suppose our supplemented description determines *S*'s content. Then, although our theory does not employ intentional *notions*, it cites sufficient conditions for mental states to have intentional *properties*. Our supposed alternative to intentional explanation actually employs theoretical resources upon which intentional explanation supervenes. At best, this is a pyrrhic victory for those who seek to avoid intentional explanation.[6]

Another idea would be to individuate syntactic types partly through relations to the environment. Although most philosophers accept *syntactic internalism*, Bontly (1998), Horowitz (2007), and Piccinini (2008, pp. 219-223) pursue an externalist alternative, according to which environment-involving properties help determine a computational system's syntactic properties. For instance, such a view might maintain that John is syntactically distinct from his twin on Twin Earth. If syntactic properties are environment-involving, then surely the domain of T-O laws conforming to (2) vastly expands.

This suggestion also faces a dilemma. Either syntactic type does not determine content, or it does. If it does not, then why should uniform environment-involving consequences ensue when a thinker bears some relation to some Mentalese sentence? If syntactic type does determine content, then SEMANTIC BARRENNESS is false, so we must abandon FCP.

§3.3 Scientific psychology requires intentional laws

I conclude that purely syntactic, non-intentional psychological theories fail to explain bodily motions under desired environment-involving descriptions. Thus, we should reject the syntactic theory of mind. My argument generalizes to show that purely neurological, non-intentional theories cannot explain bodily motions under desired descriptions. Thus, we should reject the thesis that scientific psychology can restrict itself to neuroscience, assuming that we construe neuroscience as non-intentional.

A natural further inference is that we should isolate laws describing how mental states with certain intentional properties tend to induce bodily motions with certain environment-involving properties. In other words, we should accept ILT. The argument here is by elimination: how else can we achieve our explanatory goals? This reasoning is highly defeasible. Still, it shifts the burden of proof to those who oppose intentional laws.

Philosophers sometimes argue that explaining bodily motions under appropriate environment-involving descriptions requires an externalist individuation of propositional attitudes (Peacocke, 1993), (Wilson, 1994). There may be good arguments for this conclusion, but it does not follow from anything I have said. §3.2 critiqued laws that relate mental states under *non-semantic* descriptions to bodily motions under environment-involving descriptions, not laws that relate mental states under *narrow semantic* descriptions to bodily motions under environment-involving descriptions. My argument does not depend upon any particular conception of semantics. In particular, it does not assume that content is wide. My argument aims to establish only that psychological laws should cite *some* kind of meaning, content, or semantics beyond mere uninterpreted syntax.

One might object that my argument for ILT presupposes an outdated *covering law* model of scientific explanation, according to which explaining an event requires subsuming it under a law. Although many philosophers of psychology retain something like this model, it finds numerous opponents within both philosophy of psychology and philosophy of science more generally (Cummins, 2000), (Ruben, 2003), (Schiffer, 1991), (Woodward, 2003), (Wright and Bechtel, 2007). Does my discussion rest on a defective conception of psychological explanation? I want to examine this worry in some detail.

Cummins (2000) argues that most laws isolated by scientific psychology, such as the Law of Effect, describe empirically observed regularities. These laws are explananda rather than explanantia, because mere observed regularities lack explanatory force. According to Cummins, psychological explanation involves not subsumption under laws but *functional analysis*, which "consists in analyzing a disposition into a number of less problematic dispositions such that programmed manifestation of these analyzing dispositions amounts to a manifestation of the analyzed disposition. By 'programmed' here, I simply mean organized in a way that could be specified in a program or a flow chart" (2000, p. 125). He cites computational psychology as an example of functional analysis.

Although Cummins officially opposes the covering law model, his exposition reveals that functional analysis requires laws *in my sense*. He urges that we should explicate dispositional properties through counterfactual conditionals, and he concludes: "to have a dispositional property is to satisfy a law in situ, a law characterizing the behavior of a certain kind of thing" (p. 125). Clearly, then, functional analysis posits counterfactual-supporting generalizations. Cummins denies that scientific psychology employs *general laws of nature*, which "govern all nature generally," but not that it employs *laws in situ*, which "hold of a special kind of system

because of its peculiar constitution and organization" (p. 121). We need not ponder how exactly "general laws of nature" differ from "laws in situ." The key point is that Cummins embraces laws *in my sense*. Our question therefore persists: should the laws cite intentional properties? In Cummins's jargon: should our functional analyses of cognitive systems cite dispositions to respond in certain ways to mental states with certain intentional properties?

A similar dialectic arises regarding Wright and Bechtel (2007), according to whom psychological explanation proceeds by isolating a "mechanism" that produces the explanandum. This account undercuts a nomological approach to psychological explanation only if we can describe mechanisms without articulating laws. Yet Woodward (2002) argues that any adequate description of a mechanism will prominently showcase counterfactual-supporting generalizations describing how the mechanism's components behave. Such generalizations are laws in my sense (albeit not in Woodward's sense). If Woodward's analysis is correct, then Bechtel and Wright's account is an instance of the covering law model, not a rival to it.

In response to such worries, Bechtel and Wright concede that a mechanistic theory may involve laws, but they urge that laws will play a fairly minor explanatory role (p. 54). Specifically, they argue that a suitable theory may employ graphical representations, such as diagrams, rather than linguistic representations (p. 52). However, if a diagram depicting some psychological mechanism has representational content, we can presumably express that content through a sentence, and the sentence will presumably support counterfactuals. Thus, it is difficult to see how the proposed mechanistic account undercuts the main assumption underlying my argument for ILT: that psychological explanation isolates patterns of counterfactual dependence among relevant features of the cognitive system. The question is simply whether to include intentional properties among the relevant features.

Our discussion illustrates that it is not so easy to avoid psychological laws, in my minimal sense of "law." Many accounts officially advertised as non-nomological actually require counterfactual-supporting generalizations. I conjecture that proponents and opponents of psychological laws often talk at cross-purposes, with the former employing a much weaker conception of "law" than the latter.

Ruben (2003, pp. 185-217) and Schiffer (1991) espouse a view of psychological explanation that eschews laws even in my minimal sense. They argue that folk psychology explains bodily movements through singular causal claims. For instance, we might say that John grasped a peanut because he wanted to eat it. This explanation does not explicitly cite laws. If Ruben and Schiffer are correct, then my argument for ILT seemingly collapses.

Explanation is a flexible concept. Maybe there is a sense in which folk psychology explains behavior through singular causal claims. If so, we can reformulate my argument as follows. Scientific psychology should *improve* on folk psychological explanation. In particular, scientific psychology should improve on how folk psychology explains bodily motions under appropriate environment-involving descriptions. Neither purely syntactic nor purely neurological explanation advances this goal. The only evident way to advance it is to isolate laws that subsume mental states under intentional descriptions.

Thus reformulated, my argument does not presuppose a covering law model of explanation. Consider Woodward's proposal that "explanation is a matter of exhibiting systematic patterns of counterfactual dependence" (2003, p. 191). According to Woodward, a singular causal claim of the form

Event *A* caused event *B*

is explanatory, because it supports the counterfactual

B would not have occurred if *A* had not occurred.

The singular causal claim is explanatory even if we do not associate it with any relevant law that subsumes *A* and *B*. Thus, Woodward rejects the covering law model. Yet he emphasizes that a singular causal explanation, taken on its own, is only minimally explanatory, because it does not elucidate how *B*'s properties would have been different had *A*'s been different (p. 217). To achieve a more satisfying explanation, we require a counterfactual-supporting generalization that subsumes *A* and *B* under appropriate descriptions (p. 203, p. 217, pp. 279-285). Applying Woodward's framework to bodily motion, I submit that the singular causal claim "John's desire to eat a peanut caused him to grasp a peanut," taken on its own, fails to support counterfactuals about how environment-involving features of John's behavior would have been different had various properties of his mental states been different. At best, then, the singular causal claim is minimally explanatory. To achieve a more satisfying explanation, we must isolate patterns of counterfactual dependence between environment-involving properties of bodily motions and relevant properties of mental states. (Cf. Peacocke 1993, 1994.) Given §3.2, the "relevant properties" should include intentional properties. ILT ensues.

Assuming that propositional attitudes are relations to Mentalese symbols, there are two promising ways we might try to isolate intentional laws. First, we might delineate laws that cite semantic properties of mental states without citing their syntactic properties. Second, we might delineate laws that cite both syntactic *and* semantic properties of mental states. On the first approach, espoused by Fodor (1994), Pylyshyn (1984), Schneider (2005), and many others, intentional psychology attends solely to content, ignoring the symbols that serve as vehicles for content. On the second approach, espoused by Devitt (1991), Horowitz (2007), and Perry and Israel (1991), intentional psychology alludes to both Mentalese syntax and Mentalese semantics.

I will argue that, assuming FCP, neither approach is promising. I address semantic, non-syntactic laws in §§4-7 and mixed syntactic-semantic laws in §8.

§4. Purely semantic, non-syntactic laws

Advocates of semantic, non-syntactic laws typically motivate their position by observing that computational systems heterogeneous under syntactic description may be homogeneous under semantic description. Consider Turing machines satisfying the following semantic description: if we input a symbol denoting the number *n*, then the machine outputs a symbol denoting the number 2*n*. Machines satisfying this description may vary considerably at the syntactic level. For instance, one machine may represent numbers through stroke notation, another through Arabic decimal notation. Thus, semantic description offers a different order of generality than syntactic description.

Block (1986), Cummins (1989, pp. 132-133), Fodor (1991a, 1994), and Pylyshyn (1984, pp. 23-86) argue that cognition features an analogous disparity between syntactic and intentional description. As Fodor puts it, "we *need* an intentional taxonomy because the computational one slices mental states too thin" (Fodor, 1991a, p. 311). Fodor proposes the following picture. Intentional psychology isolates *ceteris paribus* laws that cite the contents of attitudes but not the Mentalese symbols expressing those contents. Intentional laws are *implemented* by computational mechanisms sensitive to Mentalese syntax but not semantics. Systems heterogeneous under syntactic description may be homogeneous under intentional description. Computational and intentional psychology occupy distinct tiers of scientific explanation, each offering its own useful level of generality.

I devote the next two sections to critiquing Fodor's account.

§5. Fodor on Frege cases

According to Fodor, content is "Russellian": it is determined by reference, without any role for Fregean "mode of presentation" (MOP) or "way of thinking about a referent." For instance, the belief that Hesperus is large and the belief that Phosphorus is large have the same content. At the syntactic level, however, Mentalese symbols function as MOPs. Because one grasps a Russellian proposition only by grasping a Mentalese sentence that expresses it, one can bear conflicting attitudes to the same Russellian proposition (e.g. the single proposition expressed by "Hesperus is large" and "Phosphorus is large"). In this way, Fodor accommodates the Fregean "paradoxes of identity" without treating MOP as an ingredient of content.

Nevertheless, "Frege cases" pose a serious difficulty for Fodor. If we individuate content in a Russellian manner, then intentional laws do not mention MOP. A typical law might describe what consequence ensues when a thinker *J* bears certain attitudes to certain Russellian propositions.[7] *J* may grasp a Russellian proposition under distinct MOPs, i.e. as mediated by distinct Mentalese sentences *R* and *S*. By FCP, mental processes are not sensitive to semantics. In particular, they are not sensitive to the fact that *R* and *S* express the same proposition. But then what ensures that *R* and *S* occupy similar roles within *J*'s cognition? As Fodor puts it, "*Fb* believers don't, in general, behave like *Fa* believers whenever *a*=*b*" (1994, p. 23). Why should any consequence uniformly ensue when *J* bears certain attitudes to certain propositions? In principle, Fregean counter-examples may confront any putative law of Russellian intentional psychology, even if we restrict the law to a single thinker at a single instant.

Fodor responds that, although Frege cases may arise, we should treat them not as *counter-examples* to intentional laws but as *exceptions*. In his words, "intentional psychology is a special (i.e. nonbasic) science, so its laws are ceteris paribus laws. And *ceteris paribus laws*

tolerate exceptions, so long as the exceptions are unsystematic" (1994, p. 39). Quantum anomalies do not disconfirm the chemical law that water freezes at 0° Celsius. Head injuries do not disconfirm intentional laws. In both cases, other things are not equal. According to Fodor, Frege cases are likewise instances where other things are not equal.

Fodor develops this idea by observing that thinkers typically accumulate information germane to the success of their actions. If it would be useful to know whether *p*, then one generally tries to determine whether *p*. Fodor concludes that cognitive creatures normally attain *epistemic equilibrium*, in which they know what they must about the objects upon which they act. In particular, "Smith can be relied upon to know that *a=b* if the fact that *a=b* is germane to the success of his behavior" (1994, p. 41). Assuming that Smith believes *a=b*, and assuming minimal rationality, Smith also believes that *Fa* iff *Fb*. Thus, if Smith believes that *a=b*, then the belief that *Fa* and the belief that *Fb* occupy roughly the same functional role in Smith's cognition. More generally, if Smith believes that *a=b*, then *a*-thoughts occupy roughly the same functional role for Smith as comparable *b*-thoughts. The pursuit of epistemic equilibrium sustains *harmony* between mind and world, in which the mind grasps that relevant co-referential Mentalese words co-refer. Harmony undergirds laws framed at the referential level.

Epistemic equilibrium may lapse. Nevertheless, Fodor insists, lapses are *rare* and *exceptional*. Something goes wrong, because the thinker fails to discover pertinent information. To explain what goes wrong, we must descend from the intentional to the computational level, just as explaining what goes wrong during a head injury may require us to descend from the intentional to the neurological level. For instance, to explain why Hesperus-thoughts and Phosphorus-thoughts induce different behavior in ancient astronomers, we must examine the different computational roles those thoughts occupy.

§6. Criticism of Fodor on Frege cases

Rupert (2008) and Schneider (2005) follow Fodor in treating Frege cases as exceptions to intentional laws. Braun (2000) advocates a parallel position regarding folk psychological generalizations. But most commentators reject Fodor's approach.

Perry (1991) and Wakefield (2002) object that Frege cases are not rare. Admittedly, there are relatively few co-referring proper names in natural language. But not all MOPs correspond to proper names. Frege cases also arise at the predicative level (e.g. *quicksilver* versus *mercury*). Even if we restrict attention to singular terms, demonstrative thought furnishes countless examples. As Perry argues, I might think about some object under a demonstrative and a non-demonstrative MOP without recognizing that the two MOPs present the same object (e.g. *That man* versus *Jerry Fodor*). Perry's famous aircraft carrier example illustrates that I can simultaneously entertain two co-referring demonstrative MOPs without realizing that they co-refer (Perry, 1977). Since demonstrative thought pervades daily life, we should not treat Frege cases as deviations from a more pervasive epistemic equilibrium.

A second common criticism is that, whether or not Frege cases are rare, they fall under the purview of intentional psychology (Arjo, 1996), (Aydede and Robbins, 2001), (Rives, forthcoming), (Wakefield, 2002). If a baseball hits John's head, we may not expect intentional explanations for his deviant behavior. If John merely fails to believe the pertinent fact that $a=b$, we still expect an explanation that reveals how the contents of his *a*-thoughts and *b*-thoughts inform his thoughts and actions. Most pointedly, Wakefield (2002, p. 129) notes that we expect an intentional theory of how John discovers that $a=b$. For instance, we expect an intentional explanation for how ancient astronomers discovered that Hesperus is Phosphorus.

In response to this second criticism, Schneider (2005) concedes that intuition favors placing Frege cases within the ambit of intentional psychology, but she asks why we should take intuition as our guide. As long as we accommodate Frege cases somewhere within our overall theory, who cares whether we accommodate them at the computational rather than the intentional level? What do we sacrifice, beyond mere intuitions about the scope of intentional psychology? Rives (forthcoming) replies that the burden of proof lies with Fodor to isolate a principled reason for placing Frege cases outside the scope of intentional explanation. However, Fodor's original discussion already suggests a natural answer to Rives's burden-shifting argument: namely, that agents who deviate from epistemic equilibrium are somehow *defective*. I therefore want to pursue a different reply to Schneider's challenge.

I propose that we answer Schneider by incorporating the considerations from §3 into the following schematic argument:

(3) We should explain actions performed during Frege cases under appropriate environment-involving descriptions.

(4) To explain the actions under the desired descriptions, we must subsume them under laws that cite the contents of relevant mental states.

(5) Fodor cannot subsume Frege cases under relevant intentional laws, because he treats them as exceptions to intentional laws.

(6) Therefore, Fodor cannot explain the actions under the desired descriptions.

I motivated (4) in §3. I now complete the argument by motivating (3) and (5).

To motivate (3), imagine an astronomer who wonders whether Hesperus is Phosphorus. He investigates by: asking whether Hesperus is Phosphorus; staring at the night sky, tracking the motions of celestial objects; studying astronomical records; and so on. The description "asking

whether Hesperus is Phosphorus" is environment-involving. Under different circumstances, the same acoustic production might have had a different meaning. Similarly for "stares at a celestial object in the night sky." We can imagine deviant scenarios where a neurophysiological twin is located inside a planetarium or even outside during the daytime. In general, almost any description we would ordinarily apply to the astronomer's activity is environment-involving. Note that these descriptions, while *environment*-involving, are not *content*-involving. They do not mention semantic properties such as reference, truth-conditions, or even narrow content. Thus, (3)'s appeal to these descriptions is not question-begging when deployed against Fodor.

Since we have rejected the autonomy principle, I assume that scientific psychology generally tries to explain actions under appropriate environment-involving descriptions. For instance, if the astronomer inhabited a qualitatively identical world in which the morning star was not the evening star, we would surely try to explain his actions under descriptions like "stares at a celestial object in the night sky." Why should the fact that the morning star *is* the evening star induce us to abandon this explanatory project? What principled reason can Fodor cite for disavowing such a fundamental theoretical goal? Certainly, the astronomer's failure to achieve "epistemic equilibrium" provides no such reason. On the contrary, the astronomer acts precisely because he wants to determine whether he is in epistemic equilibrium. He performs environment-involving investigations to determine if Hesperus is Phosphorus.

Turn now to (5). The astronomer's inquiry begins with the state of wondering whether Hesperus is Phosphorus, which at the referential level is type-identical to the state of wondering whether Hesperus is Hesperus. No plausible T-O law describes that latter state as initiating non-trivial scientific inquiries. Apparently, then, Fodor must treat the astronomer's behavior as an *exception* to intentional laws rather than a manifestation of them. The astronomer is not in

epistemic equilibrium regarding Hesperus, so Fodor mandates a non-intentional account. He allows us to subsume the astronomer's actions only under syntactic or neurological laws.

As we saw in §3, environment-independent syntactic or neurological laws do not help explain bodily motions under environment-involving descriptions, even if we stipulate relevant semantic properties of Mentalese and environment-involving properties of bodily motions. What about environment-involving syntactic or neurological laws? These also seem unlikely to help. How can there be a law, even a *ceteris paribus* law, that relates mental states under syntactic or neurological description to bodily motions under a description like "stares at a celestial object in the night sky" or "asks if Hesperus is Phosphorus"? Whether certain mental states tend to yield these consequences depends on the *contents* of the states, not merely their syntactic or neurological properties. For instance, the appearance of Mentalese sentence "*H*=*P*" in John's "conjecture box" may tend to yield these behavioral consequences if "*H*" means Hesperus and "*P*" Phosphorus, but not if "*H*" means Mark Twain and "*P*" Samuel Clemens. This example illustrates the general conclusion of §3: psychological laws that ignore content leave many relevant properties of bodily motion unexplained.

Hence, despite official enthusiasm for intentional psychology, Fodor encounters the same objection as Stich: he fails to explain certain actions under desired environment-involving descriptions. Note that my argument centers exclusively upon *intrapersonal* Frege cases. I have considered a single subject who entertains a given Russellian proposition under distinct modes of presentation, and I have argued that Fodor's approach does not satisfactorily explain this subject's actions. In contrast with discussions such (Aydede, 1998), my argument does not depend on worries regarding type-identification of Mentalese symbols across subjects.

To rebut my argument, one might propose the following descriptivist gambit. Besides wondering whether Hesperus is Phosphorus, the astronomer also wonders whether the *F* is the *G*, for appropriate predicates *F* and *G*. For instance, *F* and *G* might respectively be *object appearing in the morning sky on certain occasions* and *object appearing in the evening sky on certain occasions*. Or *F* and *G* might respectively be *object called "Hesperus"* and *object called "Phosphorus"* (or equivalent predicates citing words in the astronomer's native language). Unlike the Russellian proposition *Hesperus = Phosphorus*, the Russellian proposition *The F = the G* is not trivial. So there might be intentional laws describing how states type-identified by this latter proposition trigger a non-trivial series of astronomical investigations.

The descriptivist gambit faces a familiar difficulty. A thinker who entertains Hesperus-thoughts need not associate Hesperus with a definite description denoting Hesperus. More generally, there is no reason why a thinker who bears some attitude to the proposition *a=b* must bear that same attitude to a corresponding proposition of the form *The F = the G*. This point has been thoroughly elaborated over the past several decades (Burge, 2005, pp. 40-43, pp. 50-51, pp. 234-235), (Donnellan, 1970), (Kripke, 1980), (Evans, 1982). Fodor (1998) develops an analogous point concerning predicates: there is no reason why a thinker must associate a predicate with non-circular necessary and sufficient conditions.

Even ignoring this objection, the descriptivist gambit fails. Grant that any thinker who bears some attitude to *a=b* also bears that attitude to a corresponding Russellian proposition *The F = the G* (call the proposition "*p*"). By Fodor's own lights, one might entertain *p* under distinct MOPs, i.e. distinct Mentalese sentences *S* and *T*. Since mental processes are not sensitive to content, *S* and *T* need not occupy similar functional roles. Thus, any putative law of Russellian psychology defined over *p* can encounter Fregean counter-examples. When such counter-

examples arise, we will find ourselves unable to explain relevant actions under desired environment-involving descriptions. Nor can we avoid the difficulty by citing a new proposition *q* that mentions further properties of the *F* and the *G*, because *q* will also encounter Frege cases. To halt the regress, one might insist that certain referents can be presented only under a single MOP. But this seems a bit desperate (Burge, 2003, pp. 304-306), (Evans, 1982, p. 16).

Rupert (2008) offers an inferentialist variant upon the descriptivist gambit. He proposes that each MOP belongs to an "inferential network" comprising those Mentalese expressions that tend to cause it and those that tend to be caused by it. In Rupert's example, the Superman MOP tends to be caused by a Mentalese sentence with the content *That guy has a red cape and blue tights*, while the Clark Kent MOP tends to be caused by Mentalese sentences with different contents. Rupert aims to explain actions performed during Frege cases by citing relevant inferential networks. To preserve Fodor's rigid segregation between intentional and syntactic description, Rupert individuates inferential networks solely in terms of the Russellian contents expressed by their component MOPs, and he does not allow intentional explanation to mention specific MOPs. Thus, a typical intentional law might have the form: if there are MOPs $x_1 \ldots, x_n$ such that a thinker bears certain relations to $x_1 \ldots, x_n$, and $x_1 \ldots, x_n$ have certain Russellian contents, and $x_1 \ldots, x_n$ belong to inferential networks whose component elements have certain Russellian contents, then certain consequences will ensue. Such a law contains quantifiers ranging over MOPs but does not mention any specific MOP.

Setting aside whether Rupert's inferentialist variant avoids our first objection to the descriptivist gambit, it does not avoid the second. Two inferential networks type-identical at the Russellian level may be type-distinct at the syntactic level. Given FCP, there is no reason why two such inferential networks should have even roughly similar behavioral consequences. Just

like Russellian propositions, Rupert's inferential networks can encounter Fregean counter-examples. When such counter-examples arise, we will find ourselves unable to explain relevant actions under desired environment-involving descriptions.

§7. Finer-grained conceptions of intentional psychology

A natural diagnosis is that the foregoing difficulties stem from Fodor's Russellian view of content. By extruding MOP from intentional individuation, Fodor renders content too coarse-grained for psychological explanation. As Arjo notes (1996, p. 244), one might hope to improve on Fodor by adopting a finer-grained taxonomization, thereby differentiating Hesperus-thoughts and Phosphorus-thoughts at the intentional level.

I think that this diagnosis is mistaken. The problems facing Fodor stem from FCP, not from Russellian intentional psychology. Consider the following widely accepted doctrine:

> **SYNTACTIC PLASTICITY:** A thinker might entertain Mentalese symbol tokens that have distinct syntactic types but express the same content.

Given SYNTACTIC PLASTICITY, the difficulties facing Fodor resurface:

(7) By RTM and SYNTACTIC PLASTICITY, two of a thinker's propositional attitudes may be type-identical at the level of content but type-distinct at the level of Mentalese syntax.

(8) By FCP, mental processes are not sensitive to the fact that the two attitudes have the same content.

(9) So the two attitudes need not occupy even roughly similar functional roles.

(10) Hence, counter-examples can arise to any putative law defined solely over the intentional contents of propositional attitudes, even if we restrict the law to a single thinker at a single instant.

Echoing Fodor, one might urge that (10) is harmless, since computational rather than intentional psychology can handle "exceptional" cases. We already rejected this maneuver, as applied to Russellian content, in §6. It seems no more palatable when applied to non-Russellian content. Actions performed during "exceptional" cases have environment-involving properties that we should explain by subsumption under intentional laws.

The inferences from (7) to (8) and from (9) to (10) seem unassailable. Can we block the inference from (8) to (9)? Following Block (1986), we might regard contents as functional roles: to have some content is to occupy some role in a thinker's cognitive economy. Thus, intentional identity guarantees sameness of functional role, even though mental processes are not sensitive to intentional identity. Two symbols with the same content fall under whichever intentional laws articulate the relevant functional role.

For certain laws, the proposal is plausible. For instance, grasping the indicative conditional may constitutively require conformance to *modus ponens*. But we cannot plausibly treat all intentional laws as "meaning-constitutive." (Cf. Fodor and Lepore, 1992, p. 183.) Many folk psychological laws, such as "The moon looks larger on the horizon," do not seem essential to the contents they cite. The proposal becomes even less credible when we consider scientific psychology, which seeks to discover intentional laws through empirical research rather than *a priori* reflection. Surely our best scientific psychology will include many intentional laws not constitutive of the contents they cite. For such laws, the inference to (9) prevails. I conclude that, once we accept SYNTACTIC PLASTICITY, a restricted version of (10) ensues.[8]

Can we reject SYNTACTIC PLASTICITY? Note that, contrary to what commentators frequently suggest, the issue here is orthogonal to the debate over content externalism. Internalist theories often conform to SYNTACTIC PLASTICITY. For instance, as Aydede (1997)

emphasizes, Fodor's (1987) "mapping" theory of narrow content (a theory subsequently abandoned by Fodor) does not distinguish between Hesperus-thoughts and Phosphorus-thoughts. Conversely, an externalist might individuate mental contents by object-dependent MOPs (Evans, 1982). In the present context, externalism is a red herring. Our question is not whether we should individuate mental content partly by relations to the external environment. Our question is whether we should individuate mental content in such a fine-grained way that distinct syntactic types always express distinct contents. (Cf. Stich, 1991, p. 249.)

Suppose we introduce Fregean senses to serve as MOPs at the intentional level. Fodor's theory already includes MOPs at the syntactic level. What ensures alignment between the two types of MOP? Why can't a thinker grasp a single (Fregean) MOP under distinct (Mentalese) MOPs? (Fodor, 1998, pp. 15-22), (Margolis and Laurence, 2007, pp. 580-581). There are many pairs of linguistic expressions that plausibly express the same Fregean sense but different Mentalese words. Mates cases like *fortnight*/*two weeks* and *ketchup*/*catsup* provide plausible examples (Fodor, 1998, pp. 37-38).[9] So do Kripke cases like *London*/*Londres* (Fodor, 1990, p. 168). Other plausible examples: *and*/*but*, *horse*/*steed*, and *dog*/*cur*.

Even if one dismisses these examples, a more abstract worry remains: once we posit both Mentalese symbols *and* Fregean senses distinct from them, there is no principled reason why such examples cannot arise. If we accept RTM, then we accept a dichotomy between *vehicle* and *content*. We posit mental representations that mediate our relations to intentional contents. If vehicle and content are truly distinct, then what ensures alignment between the individuative schemes corresponding respectively to vehicles and contents? A Fregean approach to content is finer-grained than a Russellian one, so it will encounter fewer counter-examples. But restricting a problem's frequency is not eliminating it.

As far as I can see, the only promising response to this worry is to insist that the MOPs informing content individuation are the same as those informing syntactic individuation. Consider Fodor's theory of *concepts*, according to which a concept is a symbol-type individuated by two parameters: an MOP, construed as a Mentalese syntactic-type; and a referent (1998, p. 20, p. 28). For present purposes, we may regard a Fodorian concept as an ordered pair of the form: <Mentalese syntactic-type, referent>. Fodor himself restricts the phrase "content" to the referential level, so he would not count his "concepts" as "contents." What if we diverge from him, treating concepts as contents of Mentalese symbols? Then we ensure that SYNTACTIC PLASTICITY is false, because we build syntactic-type into the individuation of content. But we also abandon Fodor's two-tiered conception of psychology. A putative law defined over Fodorian concepts is equivalent to a law that taxonomizes mental states by syntax and reference. Thus, the proposal abandons the most basic architectural feature of Fodor's picture: its segregation of syntax and semantics into distinct levels of psychological explanation.

Our discussion of Fregean intentional psychology highlights a general dilemma. Either we individuate contents without regard to Mentalese syntax-type, in which case RTM, FCP, and SYNTACTIC PLASTICITY promote counter-examples to any putative law that cites content but not syntax. Or else we individuate contents partly with respect to Mentalese syntax-type, in which case we simply abandon our search for such laws. The dilemma persists whether our "contents" are Fregean senses, mappings from contexts to truth-conditions, functional roles, sets of epistemically possible worlds, primary intensions, and so on. Although I have not argued conclusively that this dilemma is unavoidable, the burden falls on proponents of FCP to demonstrate otherwise. Lacking such a demonstration, I conclude that Fodor's difficulties do not stem primarily from his Russellian conception of intentional psychology. Russellianism

exacerbates matters, but it is not the principal source of difficulty. Apparently, FCP itself precludes a satisfactory scientific psychology based upon semantic, non-syntactic laws.

Fodor (1994, p. 15) sometimes suggests that the problems posed by Frege cases are not specific to FCP, since analogous problems arise if we replace computational psychology with neuroscience or physics. I disagree. If a neural state had had a different meaning or no meaning at all, then it would have born different relations to other neural states and to the external environment, so there is no reason to assume that it would have caused "the same output," unless perhaps we individuate outputs in question-beggingly non-intentional, environment-independent terms. There is no obvious reason to assume that neurally or physically type-identical mental states with different contents "play the same role in mental processes," or that mental processes are sensitive to neural or physical properties but not to intentional properties. We should no more assume such claims than we should assume that geological or biological processes are sensitive only to physical rather than to geological or biological properties. (Cf. Burge, 2007, pp. 316-333, pp. 347-348.) If mental processes are sensitive to content, then there is no clear reason to accept a neuroscientific or physicalist analogue to (8).

The thesis that mental processes are insensitive to semantics is not a piece of common sense. It is not a straightforward consequence of neuroscience, physics, or any non-controversial philosophical premise. It is a highly substantive doctrine with radical consequences for the scope of intentional psychology. By denying that minds exhibit certain patterns of counterfactual sensitivity, FCP restricts the counterfactual-supporting generalizations available for scientific psychology, thereby barring us from providing certain kinds of explanations.

§8. Mixed syntactic-semantic laws

I turn now to psychological laws that cite both syntactic and semantic properties of Mentalese symbols. To embrace such laws is to abandon a central Fodorian desideratum: that intentional psychology type-identify thinkers who are heterogeneous under syntactic description. The price may seem worth the benefit of achieving viable intentional explanations.

I will argue that, given FCP, mixed syntactic-semantic laws offer little promise for scientific psychology. I will stress a familiar worry: FCP bars semantic properties from playing a non-trivial role in laws that mention syntactic properties. As commentators often put it, syntax "screens off" semantics. Many philosophers find this worry quite intuitive (Fodor, 1994, p. 50), which may explain why few bother developing it in much detail. I want to highlight its gravity.

Presumably, our desired T-O laws describe how certain mental states under syntactic-semantic descriptions induce certain bodily motions under environment-involving descriptions. Are T-O laws of this form compatible with FCP? I set that question aside, focusing instead on T-T laws. There are three options worth considering: either T-T laws describe mental states under purely syntactic descriptions; or T-T laws describe explaining mental states under syntactic descriptions and output mental states under syntactic-semantic descriptions; or T-T laws describe both explaining states *and* output states under syntactic-semantic descriptions.

§8.1 Purely syntactic T-T laws

This approach, which Devitt (1989, p. 381) and Perry and Israel (1991, p. 178) champion, faces a serious challenge: it cites semantic properties of mental states as explanantia while leaving those properties unexplained (Fodor, 1991a, p. 298). Suppose that we explain John's bodily motion *M* by citing states with various semantic properties. Pick one such state *X*. Since we cite *X*'s semantic properties when explaining *M*, we naturally seek to explain *X* under a

semantic description. Our theory includes laws that describe certain states as inducing *X* under a syntactic description. Those laws help us explain why John bears certain relations to certain Mentalese sentences. By SEMANTIC BARRENNESS, the sentences' syntactic types do not determine their semantic properties. Thus, we do not explain *X* under a semantic description.

We can stipulate that *X* has certain semantic properties. However, as already emphasized, explaining a state under description *s* and noting that it also satisfies description *t* is not the same as explaining it under description *t*. Following Egan, we might further amend our account with semantic stipulations for relevant explaining states. By FCP, only the syntax of those states influences which states subsequently ensue. Our semantic stipulations are irrelevant, so they contribute no added explanatory force.

For some theoretical purposes, there is nothing wrong with stipulating that a mental state has certain unexplained semantic properties. But the procedure is clearly incomplete. We want to trace the chain of explanation back to perceptual inputs. Syntactic T-T laws block this endeavor at its first step, barring us from explaining mental states under semantic descriptions.

§8.2 Syntactic-semantic T-T laws with syntactic explaining states

Let us now consider putative laws that describe how mental states with certain syntactic properties induce mental states with certain syntactic and semantic properties. Say that our computational theory includes purely syntactic laws of the form:

(11) If a thinker *J* with property Ω bears appropriate relations to Mentalese sentences S_1…, S_n, then *J* will bear an appropriate relation to Mentalese sentence *R*, *ceteris paribus*.

Ω might encompass a species, or a cultural group, or a single thinker, and so on. The "appropriate relations" mentioned by (11) are computational relations such as *placing a*

Mentalese sentence in one's belief box. Given (11), our desired syntactic-semantic law presumably has the form:

(12) If a thinker *J* with property Ω bears appropriate relations to Mentalese sentences S_1…, S_n, then *J* will bear an appropriate relation to Mentalese sentence *R* with semantic properties ρ, *ceteris paribus*.

Note that (12) isolates a pattern of counterfactual dependence involving output state semantic properties, whereas Egan isolates only counterfactual patterns involving syntactic properties.

(12) finds no advocates in the current literature, and with good reason. How can syntactic manipulations of inherently meaningless Mentalese sentences reliably induce an output state with given semantic properties? This worry is especially pressing if we assume content externalism (Fodor, 1994, pp. 12-13), but it persists as long as we assume SUPERVENIENCE FAILURE. Given that syntactic properties do not determine semantic properties, we can offer a complete syntactic description of a computational system's activity without fixing its semantics. But then how can there be laws relating syntactically characterized explaining states to a semantically characterized output state? As Peacocke observes in a similar context, purely syntactic explaining states "cannot magic into existence the complex of non-syntactic relations required for an intentional state to have a certain content" (1994, p. 305).

The one recourse I see here is to choose Ω so that it includes non-semantic properties nomologically sufficient for *R* to have semantic properties ρ. Those properties might concern causal-covariational relations to the external world, qualitative aspects of experience, evolutionary history, teleological features of the cognitive system, and so on. Given a sufficiently detailed specification of Ω, (12) is true and counterfactual-supporting.

Even so, (12) does not help us explain why *J* enters into a state with semantic properties ρ, as evidenced by the fact that the supposed explanation overgeneralizes promiscuously. Through the same technique, we could "explain" why *J* comes to bear an appropriate relation to a Mentalese sentence with any arbitrary property φ (e.g. the property of *being my grandmother's favorite Mentalese sentence*). A mental symbol's having property φ plays no role in our account of why the cognitive system yields that symbol as its output. Our putative explanation treats as pure coincidence the fact that the symbol to which *J* bears an appropriate relation is also a symbol with property φ. Thus, while Ω may help explain why *R* has ρ, it does not help explain why *J* comes to bear an appropriate relation to a symbol with ρ. (An explanation of *Fa* & *Ga* is not an explanation of $(\exists x)(Fx \,\&\, Gx)$. For instance, an explanation of why John has a mole on his back combined with an explanation of why Mary shot John does not provide an explanation of why Mary shot someone with a mole on his back.)

§8.3 Syntactic-semantic T-T laws with syntactic-semantic explaining states

Let us now consider laws that describe how mental states with certain syntactic and semantic properties induce mental states with certain syntactic and semantic properties. Such laws presumably have the form:

(13) If a thinker *J* with property Ω bears appropriate relations to Mentalese sentences $S_1 \ldots, S_n$ with semantic properties π, then *J* will bear an appropriate relation to Mentalese sentence *R* with semantic properties ρ, *ceteris paribus*.

Horowitz (2007) advocates laws along the lines of (13), and he suggests that we deploy such laws to explain mental states under semantic descriptions. He also explicitly opposes FCP. Our question is whether we can reconcile FCP with (13).

There are two cases to consider, depending on whether any primitive Mentalese word figures in both S_1…, S_n and R. Suppose first that this is so. Then we partially fix the semantics of R by fixing the semantics of S_1…, S_n. Thus, an appropriate law of the form (13) is a trivial logical consequence of the purely syntactic (11). I argued in §8.1 that (11) does not help us explain why J enters into a mental state with certain semantic properties. How can we achieve explanatory progress by replacing (11) with a trivial logical consequence?

We cannot, as evidenced by the fact that the explanatory strategy overgeneralizes. Let φ be some arbitrary property. Assume that primitive Mentalese word w appears in both S_1…, S_n and R. Then (11) entails

(14) If a thinker J with property Ω bears appropriate relations to Mentalese sentences S_1…, S_n, one of whose component words w has property φ, then J will bear an appropriate relation to Mentalese sentences R, one of whose component words has φ, *ceteris paribus*.

By subsuming J under (14), can we explain why J bears an appropriate relation to a Mentalese word with φ? No. The putative explanation is defective, because it assigns only a trivial role to the property φ that it supposedly explains. The fact that a mental representation has property φ does not inform our account of why the cognitive system yields that representation as its output. φ is just "along for the ride." We do not explain J's mental states under a description that cites φ.

Turn now to our second case: no primitive Mentalese word figures in both S_1…, S_n and R. Call this a *disjoint instance* of (13), as opposed to an *overlapping instance*. In contrast with overlapping instances, disjoint instances do not follow logically from appropriate instances of (11). Might proponents of FCP nevertheless regard some disjoint instance as true, counterfactual-supporting, and explanatory? That seems doubtful. By FCP, S_1…, S_n would play the same role in mental computation if they meant something different or even if they meant nothing at all. In

other words, the semantic properties π mentioned by (13) are irrelevant to how mental computation processes S_1…, S_n. Precisely the same consequences would have ensued even if S_1…, S_n had not had properties π. Thus, FCP must treat any disjoint instance of (13) as equivalent to some appropriate instance of (12). I already argued that (12) does not help us explain mental states under semantic descriptions.

Some philosophers may challenge the alleged collapse of (13) into (12). FCP holds that computation is insensitive to semantics *as described syntactically*. Isn't such a view compatible with an *additional* level of mixed syntactic-semantic description that depicts computation as sensitive to both syntax and semantic? This mixed level of description could include disjoint instances of (13) not equivalent to instances of (12).

I set aside whether the proposed view is plausible. The key point is that it saves FCP only be destroying it. FCP entails

> **MEDIATION:** A change in meaning, content, or semantics influences how mental computation proceeds only if accompanied by an appropriate change in syntax.

MEDIATION is fundamental not just to FCP but to any view that regards mental computation as formal in *any* interesting sense. As Fodor puts it, "[i]t is central to a computational psychology that the effects of semantic identities and differences on mental processes must always be mediated by 'local' properties of mental representations" (1994, p. 107). MEDIATION underlies Dennett's famed description of the mind as a "syntactic engine" (1987, p. 62) and Haugeland's oft-quoted *Formalist's Motto*: "if you take care of the syntax, the semantics will take care of itself" (1989, p. 106). MEDIATION also follows from Fodor's "formality condition," discussed in §2. Yet MEDIATION precludes syntactic-semantic laws that assign semantics a non-trivial role. To posit such laws is to deny that syntax mediates how semantic properties influence other

semantic properties, which is to deny that syntax mediates the influence of semantics. Thus, I do not see how FCP, or any other view that purports to regard computation as "formal," can halt the collapse of (13)'s disjoint instances into instances of (12).

Note an instructive contrast with the purely semantic laws favored by Fodor. If epistemic equilibrium prevails, then a thinker's mental activity falls under Fodorian purely semantic laws. Such laws do not collapse into purely syntactic laws, because they achieve a distinctive order of generality. Thus, FCP leaves open the possibility *in principle* of purely semantic laws. Fodor's account prevails in certain idealized circumstances. The problem is that we want intentional explanation to apply even outside these idealized circumstances, that is, even when epistemic equilibrium lapses. In contrast, FCP leaves no room, *even in principle*, for mixed syntactic-semantic laws that assign semantics a non-trivial explanatory role. There are no circumstances, no matter how idealized, in which putative syntactic-semantic laws sustain adequate intentional explanations.

§9. Non-formal computation?

Fodor's account shows that formal mental processes can implement intentional psychological laws, provided that we restrict intentional psychology to cases of epistemic equilibrium. To that extent, FCP and ILT are consistent. However, Fodor bars us from explaining actions performed outside epistemic equilibrium under desired environment-involving descriptions. The difficulty generalizes. Given SYNTACTIC PLASTICITY, counter-examples analogous to Frege cases can arise. The only promising way to abandon SYNTACTIC PLASTICITY is to taxonomize mental states by both syntax and semantics. But this taxonomization allows syntax to screen off semantics, barring content from playing a non-trivial

role in T-T laws. Thus, FCP precludes adequate intentional explanations at both the semantic, non-syntactic level *and* the syntactic-semantic level. I conclude that FCP bars us from constructing an intentional psychology that achieves our explanatory goals.

In a sense, my discussion simply elaborates the widespread intuition that FCP allows no significant role for content in psychological explanation. Unlike most commentators, I have tried to offer a fairly systematic critique of strategies for combining FCP and ILT. No doubt a few strategies have slipped through the cracks.[10] Nevertheless, it seems advisable for advocates of intentional explanation to pursue alternatives to FCP. In particular, those who wish to combine intentional psychology with the computational theory of mind should pursue alternatives to the formal conception of computation. The literature already offers a few sallies in this direction (Cummins, 1989, pp. 131-145), (Block, 1990), (Horowitz, 2007), (Peacocke, 1994, 1999). Exploring these alternative conceptions of computation is a task for another occasion.

Notes

[1] Fodor regards the computational theory of mind as applying only to *certain* psychological processes (2000, pp. 1-7). The worries raised in this paper apply only to the relevant processes.

[2] I confine attention to the "classical" approach to cognitive science, on which mental activity involves mechanical, rule-governed, serial manipulation of finite, discrete syntactic entities. For discussion of how representational notions figure in connectionist theories, see (Ramsey, 2007).

[3] Stich argues that psychology should restrict itself to *autonomous behavioral descriptions*, which satisfy the following criterion: "if [the description] applies to an organism in a given setting, then it would also apply to any replica of the organism in that setting" (1983, p. 167). As Sanford (1986, p. 153) notes, this restriction is too weak for Stich's purposes, since an autonomous behavioral description may be environment-involving and may therefore flout Stich's own autonomy principle.

[4] Peacocke (1993) urges that scientific psychology should explain bodily movements under *object-dependent* descriptions, such as "picking up some particular peanut." Object-dependent descriptions play no role in my

argument. I presuppose only that we wish to explain bodily movements under environment-involving descriptions, which may or may not be object-dependent.

[5] I presuppose an individuative scheme that treats the motor gesture and the grasping of the peanut as the same event. Some philosophers espouse an alternative individuative scheme that treats the two descriptions as denoting different events. If anything, this alternative individuative scheme would strengthen my overall argument.

[6] See (Field, 2001, p. 76) for a different perspective.

[7] I ignore probabilistic laws, but one could easily adapt my argument to cover them.

[8] See (Block, 1993, pp. 47-49) for a response.

[9] But see (Burge, 2005, pp. 172-173).

[10] For instance, I have not discussed Dretske's (1988) account. For criticism of Dretske, see (Block, 1990).

Works Cited

Arjo, D. (1996). Sticking up for Oedipus: Fodor on intentional generalizations and broad content. *Mind and Language, 11*, 231-245.

Aydede, M. (1998). Fodor on concepts and Frege puzzles. *Pacific Philosophical Quarterly, 79*, 289-294.

---. (2005). Computationalism and functionalism: syntactic theory of mind revisited. In G. Irzik and G. Güzeldere (Eds.), *Turkish Studies in the History and Philosophy of Science*. Dordrecht: Springer.

Aydede, M. and Robbins, P. (2001). Are Frege cases exceptions to intentional generalizations?. *Canadian Journal of Philosophy, 31*, 1-22.

Bechtel, W. and Wright, C. (2007). Mechanisms and psychological explanation. In P. Thagard (Ed.), *Philosophy of Psychology and Cognitive Science*. New York: Elsevier.

Block, N. (1986). Advertisement for a semantics for psychology. In P. A. French, et. al. (Eds.), *Midwest Studies in Philosophy, Vol. X*. Minneapolis: University of Minnesota Press.

---. (1990). Can the mind change the world?. In G. Boolos (Ed.), *Meaning and Method: Essays in*

Honor of Hilary Putnam. Cambridge: Cambridge University Press.

---. (1993). Holism, hyper-analyticity, and hyper-compositionality. *Mind and Language, 8*, 1-27.

Bontly, T. (1998). Individualism and the nature of syntactic states. *British Journal for the Philosophy of Science, 49*, 557-574.

Braun, D. (2000). Russellianism and psychological generalizations. *Nous, 34*, 203-236.

Burge, T. (2003). Reply to Normore. In M. Hahn and B. Ramberg (Eds.), *Reflections and Replies*. Cambridge: MIT Press.

---. (2005). *Truth, Thought, and Reason*. Oxford: Oxford University Press.

---. (2007). *Foundations of Mind*. Oxford: Oxford University Press.

Cummins, R. (1989). *Meaning and Mental Representation*. Cambridge: MIT Press.

---. (2000). "How does it work?" vs. "What are the laws?" Two conceptions of psychological explanation. In F. Keil and R. Wilson (Eds.), *Explanation and Cognition*. Cambridge: MIT Press.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Oxford University Press.

Dennett, D. 1987. *The Intentional Stance*. Cambridge: MIT Press.

Devitt, M. (1989). A narrow representational theory of the mind. In S. Silvers (Ed.), *Rerepresentation*. Boston: Kluwer.

---. (1991). Why Fodor can't have it both ways. In B. Loewer and G. Rey (Eds.), *Meaning and Mind*. Cambridge: Blackwell.

Donnellan, K. (1970). Proper names and identifying descriptions. *Synthese, 21*, 335-358.

Dretske, F. (1988). *Explaining Behavior*. Cambridge: MIT Press.

Earman, J., Roberts., J., and Smith, S. (2002). *Ceteris paribus* lost. *Erkenntnis, 57*, 281-301.

Egan, F. (1995). Computation and content. *Philosophical Review, 104*, 181-203.

---. (1999). In defense of narrow mindedness. *Mind and Language, 14*, 177-194.

Evans, G. (1982). *The Varieties of Reference*. Oxford: Clarendon Press.

Field, H. (2001). *Truth and the Absence of Fact*. Oxford: Clarendon Press.

Fodor, J. (1981). "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." In *Representations*. Cambridge: MIT Press.

---. (1987). *Psychosemantics*. Cambridge: MIT Press.

---. (1990). *A Theory of Content and Other Essays*. Cambridge: MIT Press.

---. (1991a). Replies. In B. Loewer and G. Rey (Eds.), *Meaning in Mind*. Cambridge: Blackwell.

---. (1991b). You can fool some of the people all of the time, everything else being equal: hedged laws and psychological explanations. *Mind, 100*, 19-34.

---. (1994). *The Elm and the Expert*. Cambridge: MIT Press.

---. (1998). *Concepts*. Oxford: Clarendon Press.

---. (2000). *The Mind Doesn't Work That Way*. Cambridge: MIT Press.

Fodor, J. and Lepore, E. (1992). *Holism*. Cambridge: Blackwell.

Harman, G. (1999). *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press.

Haugeland, J. (1989). *Artificial Intelligence: The Very Idea*. Cambridge: MIT Press.

Hornsby, J. (1986). Physicalist thinking and conceptions of behavior." In P. Pettit and J. McDowell (Eds.), *Subject, Thought, and Context*. Oxford: Clarendon Press.

Horowitz, A. (2007). Computation, external factors, and cognitive explanations. *Philosophical Psychology, 20*, 65-80.

Kripke, S. (1980). *Naming and Necessity*. Cambridge: Harvard University Press.

Lange, M. (2002). Who's afraid of *ceteris paribus* laws? Or: how I learned to stop worrying and love them. *Erkenntnis, 57*, 407-423.

Margolis, E. and Laurence, S. 2007. The ontology of concepts --- abstract objects or mental representations?. *Nous, 41*, 561-593.

Newell, A. and Simon, H. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM, 19*, 113-126.

Peacocke, C. (1993). Externalist explanation. *Proceedings of the Aristotelian Society, 93*, 203-230.

---. (1994). Content, computation, and externalism. *Mind and Language, 9*, 303-335.

---. (1999). Computation as involving content: a response to Egan. *Mind and Language, 14*, 195-202.

Perry, J. (1977). Frege on demonstratives. *Philosophical Review, 86*, 474-497.

---. (1998). Broadening the mind. *Philosophy and Phenomenological Research, 58*, 223-231.

Perry, J. and Israel, D. (1991). Fodor and psychological explanation. In B. Loewer & G. Rey (Eds.), *Meaning and Mind*. Cambridge: Blackwell.

Piccinini, G. (2008). Computation without representation. *Philosophical Studies*, *137*, 205-241.

Putnam, H. (1988). *Representation and Reality*. Cambridge: MIT Press.

Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge: MIT Press.

Ramsey, W. (2007). *Representation Reconsidered*. Cambridge: Cambridge University Press.

Rives, B. (Forthcoming). Concept Cartesianism, concept pragmatism, and Frege cases. *Philosophical Studies*.

Ruben, D.-H. (2003). *Action and its Explanation*. Oxford: Clarendon Press.

Rupert, R. (2008). Frege's puzzle and Frege cases: defending a quasi-syntactic solution. *Cognitive Systems Research*, *9,* 76-91.

Sanford, D. (1986). Review of Stephen Stich's *From Folk Psychology to Cognitive Science*.

Philosophy and Phenomenological Research, 47, 149-154.

Schiffer, S. (1991). Ceteris paribus laws. *Mind, 100*, 1-17.

Schneider, S. (2005). Direct reference, psychological explanation, and Frege cases. *Mind and Language, 20*, 423-447.

Smith, M. (2007). Ceteris paribus conditionals and comparative normalcy. *Journal of Philosophical Logic, 36*, 97-131.

Stich, S. (1983). *From Folk Psychology to Cognitive Science*. Cambridge: MIT Press.

---. (1991). Narrow content meets fat syntax. In B. Loewer and G. Rey (Eds.), *Meaning in Mind*. Cambridge: Blackwell.

Wakefield, J. (2002). Broad versus narrow content in the explanation of action: Fodor on Frege cases. *Philosophical Psychology, 15*, 119-133.

Wilson, R. (1994). Causal depth, theoretical appropriateness, and individualism in psychology. *Philosophy of Science, 61*, 55-75.

---. (2004). *Boundaries of the Mind*. Cambridge: Cambridge University Press.

Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science, 69*, S366-S377.

---. (2003). *Making Things Happen*. Oxford: Oxford University Press.